Methode van Roberts – hidden line/hidden surface algoritme

Computergrafiek – Computer graphics

Inleiding

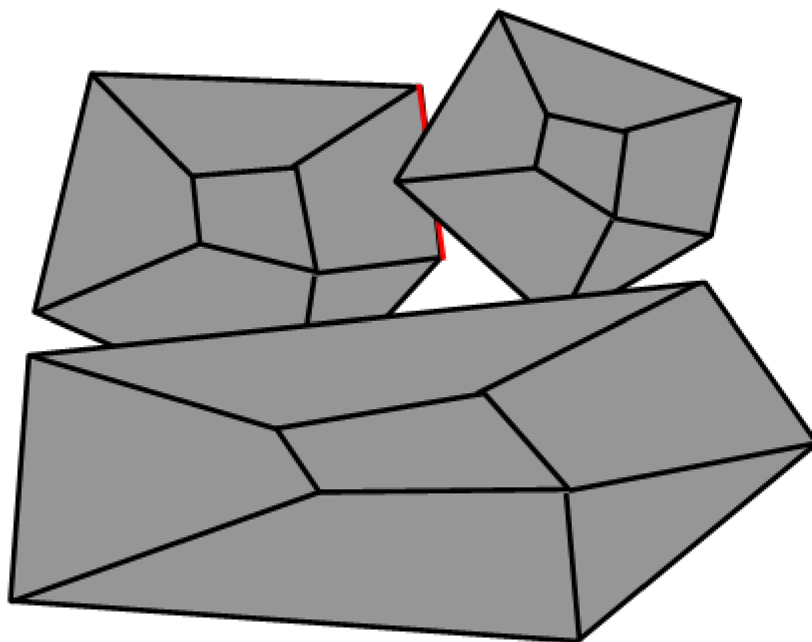
De methode van Roberts is het eerste hidden line algoritme dat gemaakt werd. Het algoritme werd in 1963 door Roberts "uitgevonden".

Doel en veronderstellingen

De bedoeling is om in een 3D ruimte te bepalen wat zichtbaar is en wat niét zichtbaar is. We veronderstellen N voorwerpen in de ruimte die elkaar niet snijden en die eveneens allemaal convexe hoeken hebben. (met andere woorden: geen enkele hoek mag > 180° zijn. Indien dit wel het geval is, kan je met een ander algoritme de ruimte eerst verdelen in kleinere voorwerpen, zodat de niet-convexe voorwerpen (ook wel concave voorwerpen) worden gesplitst en zo allemaal wél convexe voorwerpen worden).
Het Robert algoritme gaat er eveneens van uit dat al onze voorwerpen niét doorzichtig zijn.
Tot slot gaan we ervan uit dat de voorwerpen bestaan uit veelvlakken, bollen of cirkels laten we dus ook buiten beschouwing. (Ter info: je kan een cirkel of bol eventueel gaan voorstellen door zéér veel kleine driehoekjes/vierkantjes, die zo visueel hetzelfde voorwerp lijken voor te stellen)

We willen dus onze 3D ruimte gaan voorstellen op een 2D ruimte, met andere woorden een 3D ruimte op een blad of een computerscherm gaan voorstellen.

We gaan dus voor alle ribben van de voorwerpen gaan nazien of deze nu wél of niét zichtbaar zijn. (Een ribbe is de plaats waar 2 vlakken elkaar snijden.)



Doel is dus te weten dat op bovenstaande figuur het rood aangeduide wél zichtbaar is, maar een stukje niét zichtbaar is.

Onzichtbaar?

Een ribbe is onzichtbaar in 2 gevallen:

- Ze wordt bedekt door het voorwerp (veelvlak) waar ze zelf deel uit maakt. (denk aan een doos die schuin voor je staat, de achterkant kan je dan niet zien, aangezien de voorkant je het zicht belemmert.
- Ze wordt gedeeltelijk of geheel bedekt door een ander voorwerp in de ruimte.
  Het is mogelijk dat een lijnstuk in vele verschillende stukken zo visueel 'geknipt' wordt (ze zeggen dat het niet noodzakelijk enkelvoudig samenhangend is).
  Denk maar aan een borstel die voor een doos staat. Tientallen/honderden dunne balkjes (de haren van de borstel) delen de ribbe van de doos in tientallen/honderden stukjes.

We willen dus bepalen of een rechte (ribbe) nu wél of niet zichtbaar is, en dat dankzij de methode van Roberts.

Enkele definities

Alles laten we gebeuren in de zogenaamde ω-ruimte ("omega ruimte" uitgesproken).
[hiervoor heb je al wat grafische voorkennis voor nodig. De hele ruimte is bepaald in een gewoon XYZ-assenstelsel. We nemen vervolgens in die ruimte een oogpunt (punt waar we gaan kijken in de ruimte), E genoemd, en VOOR ons oog (punt E) zetten we loodrecht een scherm. Dàt scherm, een vierhoek, is de ω-ruimte. Het is dus het scherm waarop de hele 3D ruimte zal worden afgebeeld, die gezien wordt vanuit ons standpunt (het oogpunt E)]

Een punt P in onze ruimte heeft de coördinaten:

$$P = (XW, YW, ZW, W)$$

Dit zijn zogenaamde homogene coördinaten.
[Een punt (X,Y,Z) is voor te stellen in homogene coördinaten door gewoon (X, Y, Z, 1) van te maken.]

Een vlak π (pi) in onze ruimte wordt bepaald door 4 coëfficiënten en geschreven in een matrix:

$$\pi = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Verder hebben we nog een hele leuke eigenschap. Een willekeurig punt P in de ruimte ligt OP ons vlak π indien het scalair product van P en π gelijk is aan 0.
Dus:

$$P \,.\, \pi = 0$$

Of anders geschreven:

$$(XW, YW, ZW, W) \; . \; \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0$$

En deze matrixvermenigvuldiging, wordt als volgt uitgerekend en het volgende is er dus equivalent met:

$$aXW + bYW + cXW + dW = 0$$

Al het voorgaande delen door W (linker en rechterlid delen door eenzelfde van 0 verschillend getal) geeft:

$$aX + bY + cZ + d = 0$$

Dus indien aan voorgaande vergelijking voldaan, dan ligt het punt OP π, anders niet.

Indien ons punt niet op π ligt, dan ligt het er dus boven of onder. Dit maakt dus dat voorgaande formule, dan zal zijn:

$P \,.\, \pi > 0$ of $P \,.\, \pi < 0$

Verder zegt men dat het vlak π “gericht” is. Het vlak (bekijk het als een gigantisch oneindig groot blad papier dat de ruimte in 2 deelt), heeft een boven en een onderkant. Men noteert dit als een positieve en een negatieve kant.
Men kan het vlak omdraaien (dus positief en negatief omwisselen) door de coëfficiënten van het vlak te vermenigvuldigen met een bepaalde $k < 0$.

Volumematrix

We gaan verder met nog enkele definities te maken voordat we aan het algoritme zelf kunnen starten.

We definiëren een veelvlak B_i (dus een voorwerp in de ruimte) bestaande uit n_i vlakken.
Dit maakt dat:

$B_i = (\pi_{i1} \;\; \pi_{i1} \;\; \pi_{i3} \; \ldots \; \pi_{in})$

We hebben dus N voorwerpen in onze ruimte, i gaat dus van 1 tot N (ingewikkelder genoteerd: B_i ($\forall i = 1, \ldots, N$))

We gaan ons veelvlak (voorwerp) zodanig gaan definiëren dat àlle vlakken waaruit het bestaat, hun positieve kant naar BINNEN gericht zijn. Indien dit niet het geval is, wordt het vlak dus vermenigvuldigd met een $k < 0$ zodat het wel goed gericht is.
Dus IN ons voorwerp is het één en al positief, en BUITEN ons voorwerp is het langs alle kanten negatief. Dit is belangrijk, want het hele Roberts algoritme steunt hierop.

Verder hebben we ook nog onze zogenaamde “kijkbalk”. Dit is de balk die we zien. We zien namelijk op ons vlak en dat heeft een linkerkant, rechterkant, bovenkant, onderkant en een voorkant. Onze kijkbalk is half oneindig; vertrekkend vanuit E tot in het oneindige. [de benaming “halfoneindig” lijkt wel vreemd, maar dat is dus dat het ergens start, en dan in het oneindig verder gaat. Helemaal oneindig zou zijn dat het in het oneindige start, en tot in het oneindige doorloopt; wat dus duidelijk verschillend is met halfoneindig]

Met andere woorden: 5 vlakken. En deze 5 vlakken worden opgegeven in onze ω-ruimte:

$X\omega = 1$, $X\omega = -1$
$Y\omega = 1$, $Y\omega = -1$
$Z\omega = 0$

We hebben de volumematrix B_0 van de kijkpiramide. Dit is dus de matrixvoorstelling van de 5 vlakken die onze kijkbalk voorstellen:

$$B_0 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Verder gaan we de oriëntering van onze kijkbalk na of deze echt wel goed is (of dus alle 5 de vlakken wel met hun positieve kant naar binnen liggen) door het controleren dat onze Z-as (de richting waar we naar kijken) aan de positieve kant ligt.

Een bijzonder belangrijke eigenschap voor onze methode van Roberts is de volgende.
Een punt P is zichtbaar indien het IN de kijkpiramide ligt, en dus onzichtbaar als het buiten de kijkpiramide ligt. Dit controleren is doordat het punt aan de positieve zijde moet liggen van alle 5 vlakken van de kijkpiramide.
Denk hier even goed over na. Alle 5 vlakken zijn met hun positieve kant naar elkaar gericht. Als ons punt dus voor alle 5 de vlakken langs de positieve zijde ligt, ligt het dus IN de kijkpiramide, en is het dus zichtbaar.

We kunnen het voorgaande ook noteren als volgt:

$$(P \,.\, B_0) > 0$$

Dus de coördinaten van P vermenigvuldigen met de matrix van B_0, en indien die uitkomst groter is dan 0, ligt het dus aan de positieve kant van alle 5 de vlakken van onze kijkpiramide.

Het echte werk

We gaan nu met het echte werk beginnen voor de methode van Roberts.

We nemen een rechte in onze ruimte: P_1P_2. We nemen op deze rechte een generisch punt P:

$$P = \lambda P1 + (1-\lambda) P2$$
met λ in [0, 1]

Met andere woorden: het punt P ligt TUSSEN P_1 en P_2 en de positie waartussen het ligt, wordt bepaald door een parameter λ die tussen 0 en 1 ligt.

Dat punt P op onze rechte is nu ZICHTBAAR voor ons indien de coördinaten van P vermenigvuldigd met elk van de 5 vlakken van onze kijkpiramide, telkens > 0 geeft. Dus dat P langs de positieve kant ligt van alle 5 de vlakken van onze kijkpiramide.

In formulevorm geschreven:

$(\lambda P_1 + (1-\lambda) P_2) . \pi_{0i} > 0 \ , \forall i = 1, \dots 5$

Dus: $P . \pi_{0i}$ moet groter dan 0 zijn, en dat voor alle 5 de vlakken π van B_0 wat onze kijkpiramide is.

We herschrijven dit:

$\lambda ((P_1-P_2) . \pi_{0i}) > - (P_2 . \pi_{0i}) \ , \forall i = 1, \dots 5$

En indien we dit opnieuw herschrijven door het vervangen van enkele uitdrukkingen door nieuwe variabelen:

Stel:

$p_i = (P_1-P_2) . \pi_{0i}$

$t_i = - (P_2 . \pi_{0i})$

Dan:

$\lambda p_i > t_i \ , \forall i = 1, \dots 5$

Afhankelijk van het teken van p_i is dit een boven of een ondergrens voor λ.
Hier moet je ook eens even goed over nadenken om dit te begrijpen.
Indien $p_i > 0$, en we verplaatsen die naar de rechterkant van onze vergelijking, dan zegt die:

$\lambda > (t_{i)} / (p_i) \ , \forall i = 1, \dots 5$

En geeft die dus een ondergrens, want die zegt dat λ groter is dan de rechtse uitdrukking.

Indien $p_i < 0$, en we willen die naar het rechterlid verschuiven, dan dienen we de > te vervangen door een < (dat is gewoon zo volgens de wiskunde), en krijgen we dus volgende uitdrukking:

$\lambda < (t_{i)} / (p_i) \ , \forall i = 1, \dots 5$; en $(p_i < 0)$

En geeft die uitdrukking dus een bovengrens aan, want de uitkomst van de rechterkant is groter dan λ en is dus λ kleinder dan het rechterlid.

Op deze manier, door het gebruik maken van deze formules, bepalen we zo elk gedeelte van een ribbe met een functie, ook wel “parametrisch bepaald” genoemd.

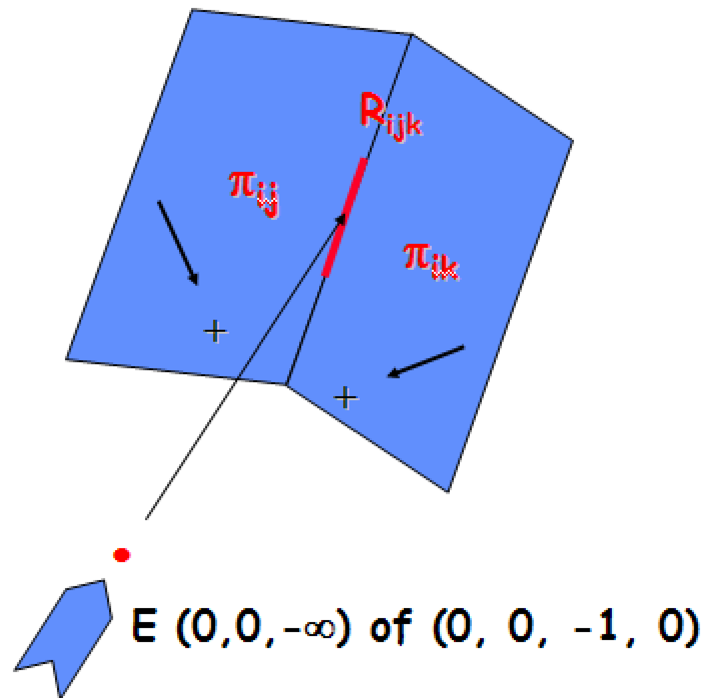
Ribbe onzichtbaar door voorwerp zelf

We nemen 2 vlakken die tegen elkaar liggen en die dus een doorsnede (intersectie) hebben. Op die intersectie hebben we een ribbe. Definieer het ene vlak als π_{ij} en het andere als π_{ik}; met dus “i” als gemeenschappelijke.
Beschouwen we nu een ribbe R_{ijk} die dus gevormd wordt door de intersectie van de twee vlakken, van het voorwerp B_i.
[Merk op: niet alle vlakken vormen 2 aan 2 een ribbe, bijvoorbeeld 2 evenwijdige vlakken hebben helemaal geen doorsnede en dus ook geen gemeenschappelijke ribbe].

Deze ribbe R_{ijk} is Onzichtbaar achter het voorwerp B_i waarvan het zelf deel uitmaakt indien het waarnemingspunt E (dus ons oog) aan de POSITIEVE kant ligt van **zowel** π_{ij} als π_{ik}.
Ah ja, natuurlijk! Als het langs beide kanten langs de positieve kant ligt, dan wil dit zeggen dat beide vlakken met hun negatieve kant weg van ons zijn (en de negatieve kant is de zichtbare kant), en dus is de ribbe bedekt door het voorwerp zélf.

Kijk maar naar een doos. Neem een ribbe die vanuit jou standpunt onzichtbaar is. Beide vlakken waaruit die bestaat, liggen allebei met hun positieve kant naar jou gericht!
Merk op dat één van beide met hun positieve kant naar jou niet genoeg is, want dan is het andere vlak nog wel met de negatieve kant naar je toe gericht én is het dus zichtbaar (want de negatieve kant naar je toegericht is nu net per definitie zichtbaar).

De volgende figuur verduidelijkt dit.



Wat je ziet op de figuur zijn dus 2 vlakken, die samen de ribbe R_{ijk} vormen. Ze zijn met hun positieve zijde naar het gezichtspunt (E) gericht, en zijn dus onzichtbaar.

Ribbe onzichtbaar door andere voorwerpen in de ruimte

We bekijken nu het tweede geval: wanneer een ribbe niet bedekt wordt door het voorwerp zelf waarvan het deel uitmaakt, maar dat het wordt gedeeltelijk of geheel onzichtbaar wordt gemaakt door àndere voorwerpen in onze ruimte.

We gaan opnieuw een generisch punt P nemen op een lijnstuk P_1P_2.
We gaan ook weer 2 vlakken beschouwen en hier ook weer een ribbe R_{pqr} op beschouwen; helemaal net zoals hiervoor.
Onze P is dus opnieuw:

$$P = \lambda P_1 + (1-\lambda) P_2$$

We willen dus de waarden van λ zoeken waarvoor het lijnstuk onzichtbaar is achter een voorwerp B_i. Die λ geeft dus de positie van P aan op het lijnstuk P_1P_2 en dus van waar tot waar het zichtbaar/onzichtbaar is. (en we gaan dat dan uiteindelijk doen voor alle voorwerpen in de hele ruimte, zodat we kunnen beslissen wat er wel/niet zichtbaar is van het lijnstuk).

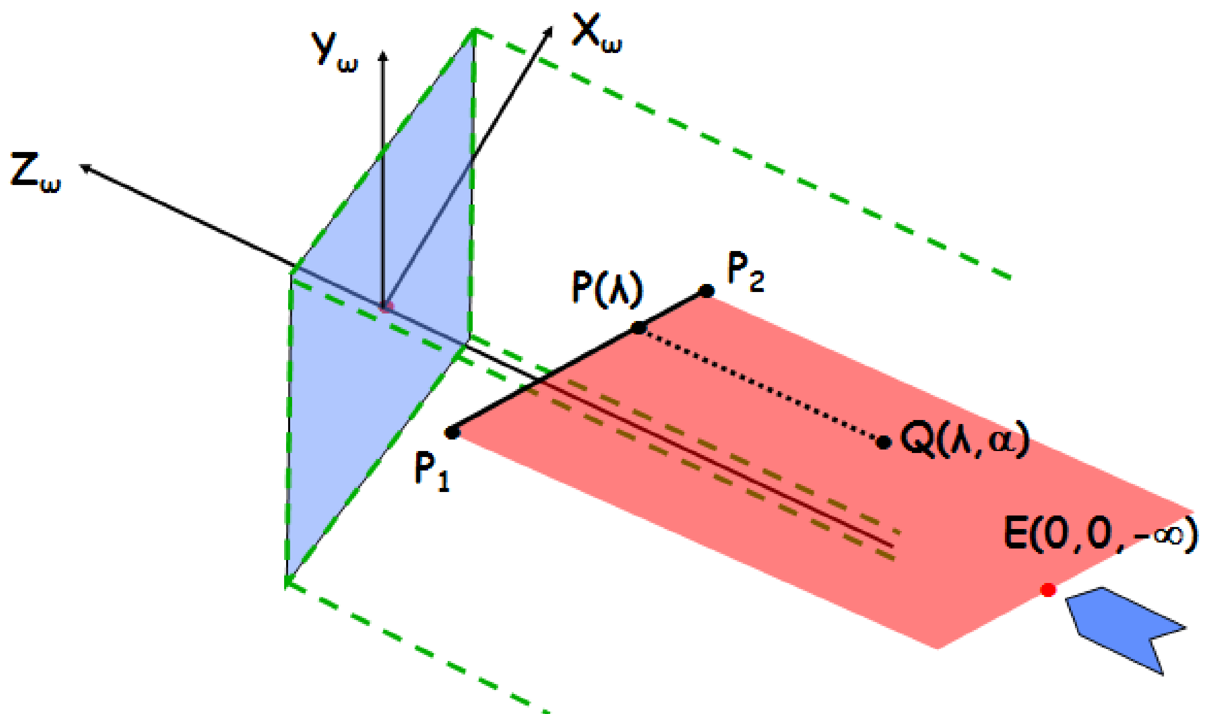
We gaan een niéuw punt beschouwen, dat we Q gaan noemen. Dit punt (in homogene coördinaten weer), wordt als volg beschouwd:

$Q = P + \alpha E$
ofwel
$Q = \lambda P_1 + (1-\lambda) P_2 + \alpha E$ met $\alpha > 0$
en $E = (0, 0, -1, 0)$ (waarnemer op $-\infty$)

Dit vraagt om een beetje extra uitleg.
Q is dus een functie van P, en eveneens van E (het oogpunt). Punt P hangt dan weer af van λ.
En dat alles waarbij de waarnemer in de homogene coördinaten $E = (0, 0, -1, 0)$ staat, of met andere woorden op $-\infty$ (min oneindig).

Voor de volledigheid, de technische definitie zegt:
"Q ligt op een halfoneindige strook rustend op P1P2 , en beschreven door een rechte die elk punt van P1P2 verbindt met E".

Een tekening maakt veel duidelijk.



Wat we zien op de tekening is dus de blauwe pijl die onze kijkrichting aangeeft met het rode bolletje, wat ons oogpunt E is. Het blauwe parallellogram is de ω-ruimte, met het assenstelsel X_ω, Y_ω en Z_ω op aangeduid. De groene stippellijnen geven de kijkpiramide aan.
We zien dan de rechte P_1P_2, met erop een punt P, dat afhankelijk is van λ, vandaar $P(\lambda)$ genoteerd.
Vanuit $P(\lambda)$ vertrekt er een rechte die naar het oogpunt E gaat, en die op een bepaalde plaats stopt, en daar ligt het punt Q, dat dus niet enkel afhankelijk van λ (want het hangt af van $P(\lambda)$), maar tevens van α, want α bepaalt de afstand ten opzichte van $P(\lambda)$.
$Q(\lambda, \alpha)$ ligt dus altijd op een rechte tussen het oogpunt E en het punt $P(\lambda)$. Het rode/roze is het vlak dat de rechte P_1P_2 en E verbindt met elkaar (en waar Q dus opligt).

En nu komt weer een idee van Roberts methode:

$P(\lambda)$ is ONzichtbaar (het ligt dus achter het voorwerp B_i) indien er een waarde voor α bestaat zodat Q een inwendig punt is van B_i.
Dit vraagt ook weer om wat extra uitleg.

Wat we eigenlijk doen is het punt $P(\lambda)$ naar ons oogpunt toekomen; en dat doen we dus met Q. We laten dat punt dus als het ware steeds dichter en dichter komen, en dat doen we door α te laten veranderen. Als we ooit een waarde voor α kunnen vinden waardoor ons punt Q BINNENIN het voorwerp B_i ligt, dàn is ons punt $P(\lambda)$ onzichtbaar. Natuurlijk, want dan wil dit zeggen dat het voorwerp B_i in de weg zit tussen ons oog en het punt $P(\lambda)$.
Neem weer de doos. En zet nog een ander voorwerp tussen jou en de doos. Neem een punt op een van de ribben, en ga in een rechte lijn vanaf dat punt naar je oog. Indien je op een bepaald ogenblik tegen dat andere voorwerp komt (en er dus normaal doorheen zou moeten gaan), dan wil dit zeggen dat het andere voorwerp dus het zicht blokkeert (in de weg staat) van ons gekozen punt en het dus onzichtbaar is!

We kunnen dit voorstellen met formules. En wel als volgt.

$$((\lambda P_1 + (1-\lambda) P_2 + \alpha E) . \pi_{ij}) > 0$$
$$\forall j = 1, \dots n_i$$

Dus het eerste deel van die functie is gewoon Q (want $Q = \lambda P_1 + (1-\lambda) P_2 + \alpha E$) en die z'n homogene coördinaten vermenigvuldigen we met de coëfficiënten van het vlak π_{ij}, en dit voor alle vlakken van ons veelvlak (voorwerp) B_i.

Indien we dus een α hebben gevonden zodat Q langs de positieve kant ligt van élk vlak waaruit het voorwerp B_i bestaat, dan ligt het dus IN dat voorwerp en is het dus niet zichtbaar voor het oog dat in E staat (want we veronderstelden dat voorwerpen niet doorzichtig zijn!).

We kunnen die formule nog eens verder herschrijven:

$$\lambda ((P_1-P_2) . \pi_{ij}) > - (P_2 . \pi_{ij}) + \alpha (E . \pi_{ij})$$
$$\forall j = 1, \dots n_i$$

Dat is dus gewoon enkele parameters van plaats wisselen en wiskundig bewerken, niets speciaals aan.

We gaan nu weer zoals we eerder hebben gedaan, enkele stukken uit deze formule vervangen door nieuwe namen, zodat het eenvoudiger lijkt.

Stel:

$$p_{ij} = (P_1-P_2) . \pi_{ij}$$
$$t_{ij} = - (P_2 . \pi_{ij})$$
$$c_{ij} = (E . \pi_{ij})$$

Zodat:

$$\lambda p_{ij} > t_{ij} + \alpha c_{ij}$$
$$\forall j = 1, \dots n_i$$

En ook hier is het weer zo: afhankelijk van het teken van p_{ij} is dit een boven of een ondergrens voor λ.
Indien $p_{ij} > 0$, dan kunnen we het dus naar de andere kant plaatsen en vormt dit:

$$\lambda > (t_{ij} + \alpha\, c_{ij}) / p_{ij}$$

en is dus λ steeds groter dan die uitdrukking rechts en is dit dus een ondergrens.

Indien $p_{ij} < 0$, dan kunnen we het dus naar de andere kant plaatsen, mits omkering van > naar < (is zo volgens de wiskunde), zodat:

$$\lambda < (t_{ij} + \alpha\, c_{ij}) / p_{ij}$$

en is dus λ steeds kleiner dan die uitdrukking rechts en is dit dus een bovengrens.

Merk op dat de waarde van deze grenzen lineair afhankelijk zijn van α. Maar ten opzichte van elkaar wijzigen ze dus niet (door het wijzigen van α wordt een bovengrens niet plots een ondergrens of zo).

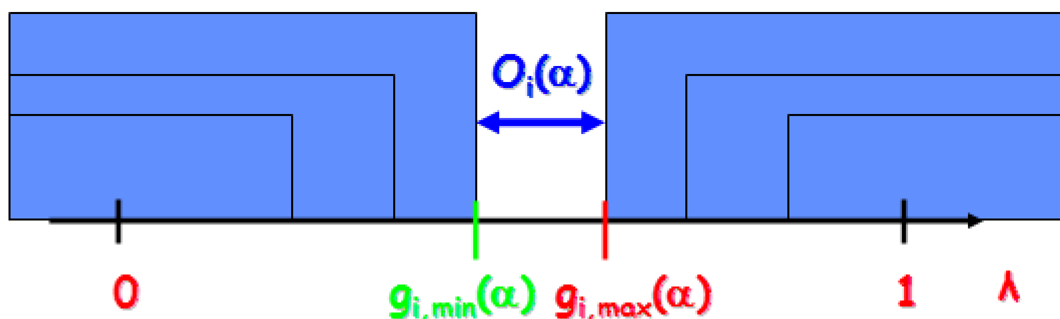
Door deze functie krijgen we een hele hoop boven en ondergrenzen. Al deze grenzen bepalen het ONzichtbaarheidsgebied voor één welbepaalde α.

Onzichtbaarheidsinterval

We definiëren weer eerst enkele zaken.

We gaan nu de grootste benedengrens nemen, en noemen deze $g_{i,min(\alpha)}$.
We benoemen ook de kleinste bovengrens, en noemen deze $g_{i,max(\alpha)}$.
$O_i(\alpha)$ is dan het interval dat afgebakend is door $g_{i,min(\alpha)}$ en $g_{i,max(\alpha)}$; en dit dus op de "diepte" α, want merk op, indien we α wijzigen (we dus dichter of verder weg gaan van ons oogpunt E), kunnen de grenzen en onzichtbaarheid verschillen.

Een figuur verduidelijkt dit ook weer extra.



Je ziet dus in het groen de grootste ondergrens, en in het rood de kleinste bovengrens voor λ.
Al de andere zijn andere grenzen die niet interessant zijn.
Het interval tussen de boven en ondergrens is dan $O_i(\alpha)$, op de figuur in het donkerblauw aangeduid.

De reden dat we dit nemen is goed te begrijpen. We nemen de meest nuttige informatie.

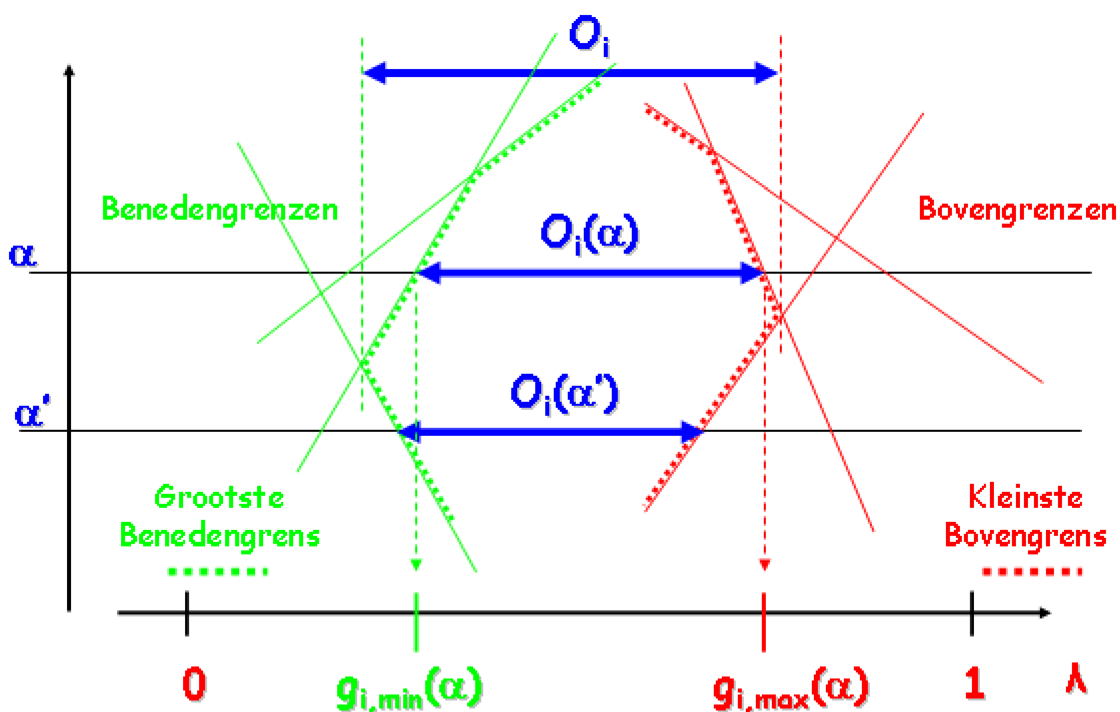
Indien we zeggen dat een getal kleiner is dan 10, kleiner is dan 15 en kleiner is dan 7; dan is het interessantste om te weten dat het kleiner is dan 7; want die andere 2 gegevens volgen hier rechtstreeks uit.
Indien men zegt dat een getal groter is dan 2, groter is dan 5 en groter is dan 3, dan is het interessantste om te weten dat het groter is dan 5, want die andere twee gegevens volgen hier ook rechtstreeks uit.
Zo weet je dan dat het getal groter is dan 5 en kleiner is dan 7; en is ons interval hiertussen.

We gaan nu O_i definiëren als volgt:

$$O_i = U_{\alpha>0}\, O_i(\alpha)$$

Of met andere woorden: O_i is de unie van alle grenzen van alle α.

Een tekening kan extra duidelijkheid brengen.



We zien op deze tekening horizontaal de as voor λ, en dat tussen 0 en 1 (dus gaande van links naar rechts op ons lijnstuk P_1P_2).
We zien α en het overeenkomstige $O_i(\alpha)$, wat dus het onzichtbaarheidsinterval is voor die welbepaalde α.
Men heeft nog een tweede uitgetekend, namelijk α', met ook weer daar zijn eigen $O_i(\alpha')$ onzichtbaarheidsinterval.
En zo zie je op de tekening steeds links in het groen allemaal puntjes wat de grootste benedengrenzen zijn, en rechts in rode puntjes de kleinste bovengrenzen.
Er zijn dus tientallen boven/ondergrenzen getekenend, enkel halen we er even α en α' uit, maar in werkelijkheid doen we dus massaal veel van dergelijke berekeningen.

O_i is dan de laatste waarde van eender welke $O_i(\alpha_i)$ en de hoogste waarde van eender welke $O_i(\alpha_i)$. En dat zie je bovenaan op de tekening.
Het is dié waarde die het uiteindelijke onzichtbaarheidsinterval is voor die welbepaalde rechte P_1P_2.

We gaan nu nog gewoon O definiëren tot slot. Deze is namelijk de unie van àlle O_i:

$$O = \bigcup_{i=1 \dots N,\, i \neq p} O_i$$

Ah ja, natuurlijk. We hadden eerst $O_i(\alpha)$, dat het onzichtbaarheidsinterval was voor één α. Dan maakten we een algemene O_i die het algemene onzichtbaarheidsinterval was voor alle α, maar wel ten opzichte van één welbepaald voorwerp B_i.
Nu maken we O, dat het meest algemene onzichtbaarheidsinterval is ten opzichte van àlle voorwerpen B, met uitzondering van B_p, want die moeten we niet meer controleren want onze rechte P_1P_2, is een ribbe van B_p en we hebben eerder al nagezien dat die onze ribbe niet onzichtbaar maakt (en dit opnieuw checken zou niet efficiënt zijn).

Deze procedure moet worden herbegonnen voor élke ribbe R_{pqr} van B_p, en voor àlle veelvlakken B_p in onze ruimte. Op deze manier kunnen we de zichtbaarheid van elke ribbe kennen ten opzichte van alle andere veelvlakken (voorwerpen) in onze ruimte en hebben we het resultaat.

Dit was het algoritme van Roberts, ofwel de Roberts methode genoemd.

Bronnen foto's:

- PowerPoint slides van Prof. Dr. Ir. Patrick P. Bergmans (Computergrafiek, Universiteit Gent, Faculteit Ingenieurswetenschappen).